

Document downloaded from:

[\[http://redivia.gva.es/handle/20.500.11939/6201\]](http://redivia.gva.es/handle/20.500.11939/6201)

This paper must be cited as:

[Cortes, V.; Barat, J. M.; Talens, P.; Blasco, J.; Jesus Lerma-Garcia, M. (2018). A comparison between NIR and ATR-FTIR spectroscopy for varietal differentiation of spanish intact almonds. Food Control, 94, 241-248.]

ivia
Institut Valencià
d'Investigacions Agràries

The final publication is available at

[\[http://dx.doi.org/10.1016/j.foodcont.2018.07.020\]](http://dx.doi.org/10.1016/j.foodcont.2018.07.020)

Copyright [Elsevier]

1 **A comparison between NIR and ATR-FTIR spectroscopy for varietal**
2 **discrimination of Spanish intact almonds**

3

4 Victoria Cortés¹, José Manuel Barat¹, Pau Talens¹, José Blasco², María Jesús Lerma-
5 García^{1,*}

6 ¹*Departamento de Tecnología de los Alimentos, Universitat Politècnica de València,*
7 *Camino de Vera s/n, 46022, Valencia, Spain*

8 ²*Centro de Agroingeniería. Instituto Valenciano de Investigaciones Agrarias (IVIA),*
9 *Ctra. CV-315, km. 10,7, 46113, Moncada, Valencia, Spain*

10

11 *Corresponding author:

12 María Jesús Lerma-García; e-mail: malerga1@tal.upv.es

13

14 **Abbreviated running title:** NIR and ATR-FTIR spectroscopy for almond varietal
15 discrimination

16

17 **ABSTRACT**

18 The rapid and easy discrimination of almond varieties with similar morphology,
19 different quality properties and in most cases different prices is interesting to protect
20 both almond industry and consumers from fraud. Therefore, in this work, intact almond
21 kernels coming from four Spanish varieties ('Guara', 'Rumbeta', 'Marcona' and
22 'Planeta') were analysed using both near infrared (NIR) and attenuated total reflectance
23 Fourier-transform infrared (ATR-FTIR) spectroscopy. After spectra measurement, an
24 attempt to classify almonds according to their variety was tried using two classification
25 methods (partial least square-discriminant analysis (PLS-DA) and quadratic

26 discriminant analysis (QDA)) applied to both NIR and ATR-FTIR pre-treated spectral
27 data. An overall accuracy of 94.45% was obtained with both PLS-DA of ATR-FTIR
28 and QDA of NIR data. These results confirm that both spectroscopic techniques, if the
29 optimal statistical model is selected, are powerful tools to reliably discriminate almonds
30 according to their varieties.

31

32 *Keywords*

33 NIR

34 ATR-FTIR

35 Intact almond

36 Varietal discrimination

37 PLS-DA

38 QDA

39

40 **1. Introduction**

41

42 Almond (*Prunus dulcis* (Mill.) D.A.Webb) is one of the main nut tree crops in
43 terms of commercial production around the world (FAOSTAT, 2012). Spain is the
44 second largest almond world producer after the US (López-Ortiz, Prats-Moya,
45 Sanahuja, Maestre-Pérez, Grané-Teruel, & Martín-Carratalá, 2008), being almond trees
46 very extended due to Spanish mild weather conditions that favour its cultivation
47 (Vázquez-Araújo, Enguix, Verdú, García-García, & Carbonell-Barrachina, 2008). In
48 addition, Spain is also an important consumer country, in which almonds are consumed
49 raw, roasted, fried or as an important ingredient in different foodstuffs like ice creams or
50 sweets as “turrón”, among others. Almond quality covers different features, such as
51 kernel and shell physical aspect and kernel organoleptic characteristics and composition
52 (with its different protein, lipid and sugar contents, among others). All these
53 characteristics are influenced by almond variety, which could define the industrial use
54 of each one of them (Cordeiro, Oliveira, Ventura, & Monteiro, 2001). There are several
55 almond varieties grown in Spain. Among them, the Marcona variety is the principal one
56 (Varela, Chen, Fiszman, & Povey, 2006), which is mainly consumed as roasted or fried
57 snack or as the main ingredient of the Protected Designations of Origin of the
58 “turrones” Jijona and Alicante. However, Marcona variety is very expensive due to their
59 excellent organoleptic properties and low production rate (Vázquez-Araújo et al., 2008).
60 Another variety to be highlighted is Guara, which has experimented and important
61 commercial triumph due to its late-flowering, self-compatibility and high quality
62 (Kodad, Estopañán, Juan, Alonso, & Espiau, 2014). Other important Spanish varieties
63 due to their large production volume are ‘Largueta’, ‘Planeta’, ‘Rumbeta’ or
64 ‘Desmayo’, among others. Therefore, it is important to find analytical methodologies

65 able to discriminate almond varieties with similar morphology or with lower prices in
66 order to protect both almond industry and consumers from fraud.

67 There are some studies published in literature that cover almond variety
68 discrimination (Gil Solsona, Boix, Ibáñez, Sancho, 2017; Piscopo, Romeo, Petrovicova,
69 & Poiana, 2010; Prats-Moya, Grané-Teruel, Berenguer-Navarro, & Martín-Carratalá,
70 1997; García-López, Grané-Teruel, Berenguer-Navarro, García-García, & Martín-
71 Carratalá, 1996), or in which almond components or physical characteristics from
72 different varieties have been established and compared (Oliveira, Meyer, Afonso,
73 Ribeiro, & Gonçalves, 2018; Zamany, Samadi, Kim, Keum, & Saini, 2017; Yada,
74 Lapsley, & Huang, 2011; Valdés, Vidal, Beltrán, Canals, & Garrigós, 2015, Kodad et
75 al., 2014; Özcan, Ünver, Erkan, & Arslan, 2011; López-Ortiz et al., 2008; Cherif, Sebei,
76 Boukhchina, Kallel, Belkacemi, & Arul, 2004; Cordeiro et al., 2001); however, the
77 analytical techniques employed are in most cases expensive, destructive and time-
78 consuming, and sample pre-treatment is normally required. Therefore, there is a need of
79 non-destructive and fast alternative methodologies able to cover this issue. In this
80 regard, the employment of spectroscopic techniques, such as infrared (IR) spectroscopy,
81 could be an excellent alternative. The potential of this technique in both, near and
82 medium IR regions, has been demonstrated in several previous works in the almond
83 field. For example, Fourier-transform infrared spectroscopy (FTIR) has been applied to
84 quality control of medicinal almonds (Chun-Song et al., 2017), while near infrared
85 spectroscopy (NIR) has been used to detect hidden damage in raw almonds (Rogel-
86 Castillo, Boulton, Opastpongkarn, Huang, & Mitchell, 2016), to inspect internal
87 damages in almonds (Nakariyakul, 2014), to discriminate sweet and bitter almonds
88 (Borrás, Amigo, van den Berg, Boqué, Busto, 2014; Cortés, Talens, Barat, & Lerma-
89 García, 2018), and to detect fungal infection in almond kernels (Liang, Slaughter,

90 Ortega-Beltran, & Michailides, 2015), among others. We have only found three articles
91 regarding almond discrimination according to their variety using IR data. Two of these
92 articles were from a research group of the University of Alicante (Beltrán Sanahuja,
93 Prats Moya, Maestre Pérez, Grané Teruel, Martín Carratalá, 2009; Beltrán, Ramos,
94 Grané, Martín, & Garrigós, 2011), in which almond varieties were discriminated after
95 almond oil extraction according to its thermal stability after application of a forced
96 oxidative treatment. For this purpose, oil degradation was studied by registering the
97 changes produced in the most abundant fatty acids (established by gas chromatography
98 (GC)) (Beltrán Sanahuja et al., 2009) or volatile compounds (established by headspace
99 solid-phase microextraction/GC–mass spectrometry (HS-SPME/GC–MS) (Beltrán et
100 al., 2011) and to changes produced in the FTIR spectra (Beltrán Sanahuja et al., 2009;
101 Beltrán et al., 2011). Using stepwise linear discriminant analysis (LDA), authors were
102 able to classify almond varieties using fatty acid contents and FTIR data in the first
103 work (Beltrán Sanahuja et al., 2009), and using HS-SPME/GC–MS data in the second
104 one (Beltrán et al., 2011). In the third article, Valdés et al. (Valdés, Beltrán, & Garrigós,
105 2013) employed FTIR and two thermal analysis techniques (differential scanning
106 calorimetry and thermogravimetric analysis) to classify almonds according to their
107 cultivar, after almond grinding and sieving. Next, LDA models were constructed
108 using FTIR and thermal data all together and separately. With these models, good
109 almond classifications according to their variety were obtained. However, as far as
110 were are concern, any article has been published regarding the employment and
111 comparison of both NIR and FTIR data to classify almonds according to their variety by
112 directly measuring spectra on intact almonds surface.

113 Therefore, the aim of this work was to explore the viability of both NIR and
114 FTIR data to reliably classify Spanish almonds according to their variety. For this

115 purpose, almonds belonging to four of the main varieties cultivated in Spain ('Guara',
116 'Rumbeta', 'Marcona' and 'Planeta') were directly measured on both spectrometers.
117 Using both, NIR and FTIR data, two different classification methods (partial least
118 square discriminant analysis (PLS-DA) and quadratic discriminant analysis (QDA))
119 were constructed and their overall accuracies compared.

120

121 **2. Materials and methods**

122

123 *2.1. Raw material*

124

125 A total of 120 almonds, coming from four different Spanish varieties ('Guara'
126 (G), 'Rumbeta' (R), 'Marcona (M) and 'Planeta' (P)), were analysed in this study. All
127 samples, gently provided by Agricoop (Alicante, Spain), were free of visual damage and
128 of uniform colour and size.

129

130 *2.2. Spectra acquisition*

131

132 *2.2.1. NIR*

133

134 An AvaSpec-NIR256-1.7 NIRLine spectrometer (AVS-DESKTOP-USB2,
135 Avantes BV, The Netherlands) was used for collecting NIR spectra of intact almond
136 kernel (with skin) over the range of 1000–1700 nm at an interval of 3.535 nm. The
137 instrument is equipped with a 10-W tungsten halogen light source (AvaLight-HAL-S,
138 Avantes BV, The Netherlands). Almond spectra were acquired in diffuse reflectance
139 mode using a bi-directional fibre-optic probe (FCR-7IR200-2-45-ME, Avantes BV, The

140 Netherlands) designed under an angle of 45° to prevent direct back-reflection from
141 almond surface. The probe, composed by two legs, is connected to the light source and
142 to the spectrometer. The integration time (500 ms) was adjusted using a 99% reflective
143 white reference (WS-2, Avantes BV, The Netherlands), so that the maximum
144 reflectance value was over 90% of saturation (Lorente, Escandell-Montero, Cubero,
145 Gómez-Sanchis, & Blasco, 2015). The dark spectrum was obtained by turning off the
146 light source and covering the tip of the reflectance probe.

147 A personal computer equipped with the commercial software AvaSoft version
148 7.2 (Avantes, Inc.) was used to acquire the spectra. For each sample, five replicates
149 were collected on both almond sides and mean spectra values were used for the
150 analysis. All measurements have been performed at room temperature (22±1 °C).

151

152 2.2.2. ATR-FTIR

153

154 ATR-FTIR spectra were obtained using a Tensor 27 spectrometer (Bruker Optics,
155 Milan, Italy) coupled to a deuterated triglycine sulphate (DTGS) detector and to an ATR
156 accessory (Specac Inc., Woodstock, Georgia, USA) composed of a zinc selenide (ZnSe)
157 crystal. Absorbance spectra were obtained in the wavenumber range from 4000 to 600
158 cm⁻¹ acquiring 32 scans per sample at a resolution of 4 cm⁻¹. After every scan, a new
159 reference air background spectrum was taken. Each intact almond kernel was put on the
160 ZnSe crystal for measurements, and the crystal was carefully cleaned by scrubbing with
161 acetone and dried with a soft tissue before measuring the next sample. The system was
162 operated using the OPUS software version 5.0 provided by Bruker Optics. Two
163 measurements were acquired for each almond (one measurement on each almond face),
164 being spectra mean employed for statistical analysis.

165

166 *2.3. Data pre-processing and multivariate analysis*

167

168 To execute the pre-treatments and multivariate procedures, ‘The Unscrambler X’
169 software version 10.3 (Camo Process SA, Trondheim, Norway) was used.

170 Before multivariate analysis, the dispersion of almond NIR spectra was
171 corrected by simultaneously applying Savitzky-Golay (S-G) smoothing (3 points gap),
172 extended multiplicative scatter correction (EMSC) and second derivative (with a 2.3
173 gap-segment). In the case of ATR-FTIR, standard normal variate (SNV) and S-G
174 second derivative (with a second order polynomial) spectral pre-treatments were applied.

175 After spectra pre-treatments, principal component analysis (PCA) models were
176 constructed to obtain qualitative information about the possible varietal discrimination
177 and to identify possible outliers.

178 In order to construct the chemometric models for both, NIR and ATR-FTIR
179 data, the full sample set (N= 120) was divided into training (70% of almonds) and test
180 sets (remaining 30% of almonds). Once the models were constructed, and before
181 external validation with the test set, model were internally validated using full cross-
182 validation (CV; leave-one-out method) (Huang, Yu, Xu, & Ying, 2008).

183 Two classification models (PLS-DA and QDA) to differentiate almond varieties
184 were constructed with both NIR and ATR-FTIR data. The PLS-DA models were
185 constructed using the PLS algorithms (Wold, Sjöström, & Eriksson, 2011), where the
186 variables in the *X*-matrix (which corresponded to the spectral data) were related to the
187 classes included in the *Y*-matrix. This matrix contained dummy variables that describe
188 the belonging of each training set sample to a given category. The *Y*- or dummy-matrix
189 is composed by 4 columns (one column for each variety) with ones and zeros, such that

190 the entry in the first column is unity and the entry of the rest of the columns is zero for
191 the samples of the first variety, and so on until completing the 4 columns. Almond
192 classification according to their variety was performed using the 0.5 cutoff value
193 (Cortés, Ortiz, Aleixos, Blasco, Cubero, & Talens, 2016). Predicted values higher than
194 0.5 indicated that the sample belongs to a given class, while values lower than 0.5
195 indicated that the samples does not belong to this category.

196 PLS-DA models accuracy was evaluated by the number of latent variables
197 (LVs), the coefficient of determination of calibration (R^2_C), the root mean square error
198 of calibration (RMSEC), the coefficient of determination for cross-validation (R^2_{CV}) and
199 the root-mean square error of cross-validation (RMSECV).

200 For QDA models, a categorical value (Y-variable) was assigned with a different
201 letter (G, M, P and R) for each variety. To construct QDA models, a number of
202 variables lower than the number of objects is required (Sádecká, Jakubíková, Májek, &
203 Kleinová, 2016). Then, a variable reduction is needed before model construction. This
204 variable reduction is performed using PCA scores, since principal components (PCs) are
205 found as linear transformations that are uncorrelated (Rodríguez-Campos, Escalona-
206 Buendía, Orozco-Avila, Lugo-Cervantes, & Jaramillo-Flores, 2011).

207 Finally, PLS-DA and QDA models performance was evaluated by considering the
208 percentage of correctly classified test samples.

209

210 **3. Results and discussion**

211

212 *3.1. Characteristics of NIR and ATR-FTIR almond spectra*

213

214 Fig. 1 represents the typical raw and pre-processed (a) NIR and (b) ATR-FTIR
215 almonds spectra. The main absorbance bands in the NIR spectra (Fig. 1a) were
216 evidenced at 1120, 1200 and 1440 nm. These bands are representative of the chemical
217 or functional groups of components present in the almonds. The 1120 and 1200 nm
218 bands denote absorptions that may occur due to the second overtone vibration of C-H
219 stretching, while the band at 1440 nm may belong to the first overtone of O-H
220 stretching of water (Workman Jr, & Weyer, 2008).

221 Fig. 1b represents the almond ATR-FTIR spectra showing the major peaks at
222 2940, 2460, 2350, 2220, 1860, 1750, 1390, 1220 and 1040 cm^{-1} . Absorbance at 2940
223 cm^{-1} is due to the asymmetric bands arising from CH_2 stretching vibrations (Sinelli,
224 Cosio, Gigliotti, & Casiraghi, 2007), whereas the peaks at 2460, 2350 and 2220 cm^{-1}
225 could be assigned to alkane stretching (Kök, Varfolomeev, & Nurgaliev, 2017). The
226 two absorption peaks at 1860 and 1750 cm^{-1} are the characteristic peaks of the C=O
227 stretching vibrations (Beltrán Sanahuja et al., 2009; Vlachos, Skopelitis, Psaroudaki,
228 Konstantinidou, Chatzilazarou, & Tegou, 2006; Zhang, Guo, & Zhang, 2002). The peak
229 at 1390 cm^{-1} may be due to CH bending (Hernández, & Zacconi, 2009), while the peak
230 at 1220 cm^{-1} could be associated with the C-O stretching vibration (Paradkar,
231 Sakhamuri, & Irudayaraj, 2002). Finally, the peak at 1040 cm^{-1} may be due to
232 combination of vibrations of C(1)H bending (that is C-H bond at C1 position) of
233 carbohydrates (Paradkar et al., 2002).

234

235 3.2. PCA analysis

236

237 Both NIR and ATR-FTIR spectra were pre-processed before PCA model
238 construction. A preliminary data exploration with PCA was carried out with the training

239 set samples. As observed in the PCA score plots (Fig. 2a,b), an evident separation of
240 almonds according to the different varieties is observed with both NIR and ATR-FTIR
241 data. The two first PCs summarized 76% and 97% accumulative contribution of the
242 original data for NIR and ATR-FTIR data, respectively, which means that nearly all the
243 variation of the variables were explained by these PCs. Next, the X-loading plots (Fig.
244 2c,d) were analysed to evidence which variables showed the greatest separation among
245 almond varieties. As observed in PC1 and PC2 X-loading plots for the NIR data (Fig.
246 2c), the most prominent peaks were observed at 1150 nm (second overtone vibration of
247 C-H stretching) (Workman Jr et al., 2008), 1490 and 1520 nm (O-H bond stretching
248 and first water overtone) (Blanco, Coello, Iturriaga, MasPOCH, & Pages, 2000), 1570 nm
249 (N-H first overtone) (Kaddour, Mondet, & Cuq, 2008) and 1610 nm (related to
250 carbohydrate content) (Teena, Manickavasagan, Ravikanth, & Jayas, 2014), while for
251 the ATR-FTIR data the most relevant peaks were those located at 2350 (alkane
252 stretching) and 1750 cm^{-1} (C=O stretching vibrations) (Kök et al., 2017; Zhang et al.,
253 2002).

254

255 *3.3. Classification of almonds according to their variety*

256

257 Two different classification techniques (PLS-DA and QDA) were applied to
258 both NIR and ATR-FTIR pre-processed spectra in order to discriminate almonds
259 according to their variety.

260 The PLS-DA models were constructed using 7 and 14 LVs for NIR and ATR-
261 FTIR spectra, respectively. The accuracy of the PLS-DA models obtained using both
262 NIR and ATR-FTIR pre-treated data with the training set samples is included in Table
263 1. As it can be observed in this table, both spectroscopic techniques provided similar

264 and good results, with R^2_{CV} and RMSECV values comprised between 0.85-0.92 and
265 0.12-0.18, respectively. When these models were validated with the test set samples,
266 satisfactory classification rates were obtained (see Table 2). The best PLS-DA model
267 which produce the highest overall rate of correct classification was obtained using ATR-
268 FTIR data, with a 94.45% of correctly classified almonds, being this value lower
269 (86.13% of overall accuracy) for the model constructed with NIR data. The same results
270 are confirmed in Fig. 3.

271 Next, QDA models using both spectroscopic techniques data were constructed
272 using the first 9 PCs. An overall rate of 100% and 96% of correct classified samples of
273 the training set samples were obtained using NIR and ATR-FTIR data, respectively.
274 The results obtained for the test set samples are shown in Table 2. As it can be observed
275 in this table for NIR data, the almonds coming from ‘Guara’ and ‘Rumbeta’
276 varieties were both 100% correctly classified, while the samples of ‘Marcona’ and
277 ‘Planeta’ varieties were both 88.9% correctly classified. In the case of ATR-FTIR data,
278 the overall accuracy classification is lower (77.8%) than those obtained using NIR
279 (94.45%). Concretely, the samples of ‘Planeta’ and ‘Rumbeta’ varieties were both
280 88.9% correctly classified, while samples of ‘Marcona’ and ‘Guara’ varieties provided a
281 77.8% and 55.6% correctly classified samples, respectively. The QDA plots obtained
282 with both NIR and ATR-FTIR data are shown in Fig. 4. The same results of Table 2 are
283 also evidenced in this figure, where there is a good classification of samples into their
284 corresponding category for the QDA model constructed with NIR data (Fig. 4a). On the
285 other hand, the QDA model constructed with the ATR-FTIR data (Fig. 4b) evidenced
286 several misclassified samples.

287 Finally, when PLS-DA and QDA models obtained using NIR and ATR-FTIR
288 data were compared, it is possible to conclude that the best results in terms of overall

289 performance were obtained using PLS-DA of ATR-FTIR and with QDA of NIR data.
290 Therefore, these results confirm that both spectroscopic techniques, if the optimal
291 statistical model is selected, are useful for almond varietal discrimination.

292

293 **4. Conclusions**

294

295 The results obtained by the two classification methods (PLS-DA and QDA)
296 applied to both NIR and ATR-FTIR pre-processed data demonstrated that, when the
297 optimal classification method was applied, it is possible to correctly discriminate
298 Spanish almonds according to their variety. Concretely, the best overall accuracies
299 (94.45%) were obtained with the PLS-DA model of ATR-FTIR and the QDA model of
300 NIR data. Therefore, both spectroscopic techniques could be successfully applied for
301 the rapid and non-destructive varietal classification of intact almonds. The developed
302 methodology could be very useful to protect both almond industry and consumers from
303 fraud, since the almond varieties studied are from similar appearance and cover
304 different price ranges in the market.

305

306 **Acknowledgements**

307

308 Victoria Cortés López thanks the Spanish Ministry of Education, Culture and
309 Sports for the FPU grant (FPU13/04202). The authors wish to thank the cooperative
310 Agricoop for kindly providing the samples used in the experiments. This work was
311 partially funded by INIA and FEDER funds through project RTA2015-00078-00-00.

312 **References**

313

314 Beltrán Sanahuja, A., Prats Moya, M.S., Maestre Pérez, S. M., Grané Teruel, N., &
315 Martín Carratalá, M.L. (2009). Classification of four almond cultivars using oil
316 degradation parameters based on FTIR and GC data. *Journal of the American Oil*
317 *Chemists' Society*, 86, 51–58.

318 Beltrán, A., Ramos, M., Grané, N., Martín, M. L., & Garrigós, M. C. (2011).
319 Monitoring the oxidation of almond oils by HS-SPME–GC–MS and ATR-FTIR:
320 Application of volatile compounds determination to cultivar authenticity. *Food*
321 *Chemistry*, 126(2), 603–609.

322 Blanco, M., Coello, J., Iturriaga, H., MasPOCH, S., & Pages, J. (2000). NIR calibration in
323 non-linear systems: different PLS approaches and artificial neural networks.
324 *Chemometrics and Intelligent Laboratory Systems*, 50(1), 75–82.

325 Borrás, E., Amigo, J. M., van den Berg, F., Boqué, R., & Busto, O. (2014). Fast and
326 robust discrimination of almonds (*Prunus amygdalus*) with respect to their
327 bitterness by using near infrared and partial least squares-discriminant analysis.
328 *Food Chemistry*, 153, 15–19.

329 Cherif, A., Sebei, K., Boukhchina, S., Kallel, H., Belkacemi, K., & Arul, J. (2004).
330 Kernel fatty acid and triacylglycerol composition for three almond cultivars
331 during maturation. *Journal of the American Oil Chemists' Society*, 81(10), 901–
332 905.

333 Chun-Song, C., Can-Jian, W., Liang, J., Chi-Chou, L., Hua, Z., Zhin-Feng, & Zhang, Z.
334 F. (2017). A new approach for identification of medicinal almonds by fourier
335 transform infrared spectroscopy and systematic clustering of characteristic peaks.
336 *Chinese Journal of Natural Medicines*, 15, 703–709.

337 Cordeiro, V., Oliveira, M., Ventura, J., & Monteiro, A. (2001). Study of some physical
338 characters and nutritive composition of the Portuguese's (local) almond varieties.
339 *Options Méditerranéennes*, 56, 331–337.

340 Cortés, V., Ortiz, C., Aleixos, N., Blasco, J., Cubero, S., & Talens, P. (2016). A new
341 internal quality index for mango and its prediction by external visible and near-
342 infrared reflection spectroscopy. *Postharvest Biology and Technology*, 118, 148–
343 158.

344 Cortés, V., Talens, P., Barat, J. M., & Lerma-García, M. J. (2018). Potential of NIR
345 spectroscopy to predict amygdalin content established by HPLC in intact almonds
346 and classification based on almond bitterness. *Food Control*, 91, 68–75.

347 FAOSTAT (2012). FAO Statistical Databases. Food and Agriculture Organisation of
348 the United Nations, Statistics Division, Available at (<http://faostat3.fao.org>).

349 García-López, C., Grané-Teruel, N., Berenguer-Navarro, V., García-García, J. E., &
350 Martín-Carratalá, M. L. (1996). Major fatty acid composition of 19 almond
351 cultivars of different origins. A chemometric approach. *Journal of Agricultural
352 and Food Chemistry*, 44(7), 1751–1755.

353 Gil Solsona, R., Boix, C., Ibáñez, M., & Sancho, J. V. (2017). The classification of
354 almonds (*Prunus dulcis*) by country and variety using UHPLC-HRMS-based
355 untargeted metabolomics. *Food Additives & Contaminants: Part A*,
356 <https://doi.org/10.1080/19440049.2017.1416679>.

357 Hernández, S. A., & Zacconi, F. (2009). Aceite de almendras dulces: Extracción,
358 caracterización y aplicación. *Química Nova*, 32(5), 1342–1345.

359 Huang, H., Yu, H., Xu, H., & Ying, Y. (2008). Near infrared spectroscopy for on/in-line
360 monitoring of quality in foods and beverages: a review. *Journal of Food
361 Engineering*, 87(3), 303–313.

362 Kaddour, A. A., Mondet, M., & Cuq, B. (2008). Application of two-dimensional cross-
363 correlation spectroscopy to analyse infrared (MIR and NIR) spectra recorded
364 during bread dough mixing. *Journal of Cereal Science*, 48(3), 678–685.

365 Kodad, O., Estopañán, G., Juan, T., Alonso, J. M., & Espiau, M. T. (2014). Oil content,
366 fatty acid composition and tocopherol concentration in the Spanish almond
367 genebank collection. *Scientia Horticulturae*, 177, 99–107.

368 Kök, M. V., Varfolomeev, M. A., & Nurgaliev, D. K. (2017). Thermal characterization
369 of crude oils in the presence of limestone matrix by TGA-DTG-FTIR. *Journal of*
370 *Petroleum Science and Engineering*, 154, 495–501.

371 Liang, P.S., Slaughter, D.C., Ortega-Beltran, A., & Michailides, T.J. (2015). Detection
372 of fungal infection in almond kernels using near-infrared reflectance
373 spectroscopy. *Biosystems Engineering*, 137, 64–72.

374 López-Ortiz, C. M., Prats-Moya, S., Sanahuja, A. B., Maestre-Pérez, S. E., Grané-
375 Teruel, N., & Martín-Carratalá, M. L. (2008). Comparative study of tocopherol
376 homologue content in four almond oil cultivars during two consecutive years.
377 *Journal of Food Composition and Analysis*, 21(2), 144–151.

378 Lorente, D., Escandell-Montero, P., Cubero, S., Gómez-Sanchis, J., & Blasco, J. (2015).
379 Visible-NIR reflectance spectroscopy and manifold learning methods applied to
380 the detection of fungal infections on citrus fruit. *Journal of Food Engineering*,
381 163, 17–21.

382 Nakariyakul, S. (2014). Internal damage inspection of almond nuts using optimal near-
383 infrared waveband selection technique. *Journal of Food Engineering*, 126, 173–
384 177.

385 Oliveira, I., Meyer, A., Afonso, S., Ribeiro, C., & Gonçalves, B. (2018). Morphological,
386 mechanical and antioxidant properties of Portuguese almond cultivars. *Journal of*
387 *Food Science and Technology*, 55(2), 467–478.

388 Özcan, M. M., Ünver, A., Erkan, E., & Arslan, D. (2011). Characteristics of some
389 almond kernel and oils. *Scientia Horticulturae*, 127(3), 330–333.

390 Paradkar, M. M., Sakhamuri, S., & Irudayaraj, J. (2002). Comparison of FTIR,
391 FT-Raman, and NIR spectroscopy in a maple syrup adulteration study. *Journal of*
392 *Food Science*, 67(6), 2009–2015.

393 Piscopo, A., Romeo, F. V., Petrovicova, B., & Poiana, M. (2010). Effect of the harvest
394 time on kernel quality of several almond varieties (*Prunus dulcis* (Mill.) DA
395 Webb). *Scientia Horticulturae*, 125(1), 41–46.

396 Prats-Moya, S., Grané-Teruel, N., Berenguer-Navarro, V., & Martín-Carratalá, M. L.
397 (1997). Inductively coupled plasma application for the classification of 19 almond
398 cultivars using inorganic element composition. *Journal of Agricultural and Food*
399 *Chemistry*, 45(6), 2093–2097.

400 Rodriguez-Campos, J., Escalona-Buendía, H. B., Orozco-Avila, I., Lugo-Cervantes, E.,
401 & Jaramillo-Flores, M. E. (2011). Dynamics of volatile and non-volatile
402 compounds in cocoa (*Theobroma cacao* L.) during fermentation and drying
403 processes using principal components analysis. *Food Research International*, 44,
404 250–258.

405 Rogel-Castillo, C., Boulton, R., Opastpongkarn, A., Huang, G., & Mitchell, A.E.
406 (2016). Use of Near-Infrared Spectroscopy and Chemometrics for the
407 Nondestructive Identification of Concealed Damage in Raw Almonds (*Prunus*
408 *dulcis*). *Journal of Agricultural and Food Chemistry*, 64(29), 5958–5962.

409 Sádecká, J., Jakubíková, M., Májek, P., & Kleinová, A. (2016). Classification of plum
410 spirit drinks by synchronous fluorescence spectroscopy. *Food Chemistry*, *196*,
411 783–790.

412 Sinelli, N., Cosio, M. S., Gigliotti, C., & Casiraghi, E. (2007). Preliminary study on
413 application of mid infrared spectroscopy for the evaluation of the virgin olive oil
414 “freshness”. *Analytica Chimica Acta*, *598*, 128–134.

415 Teena, M. A., Manickavasagan, A., Ravikanth, L., & Jayas, D. S. (2014). Near infrared
416 (NIR) hyperspectral imaging to classify fungal infected date fruits. *Journal of*
417 *Stored Products Research*, *59*, 306–313.

418 Valdés, A., Beltrán, A., & Garrigós, M. C. (2013). Characterization and classification of
419 almond cultivars by using spectroscopic and thermal techniques. *Journal of Food*
420 *Science*, *78*, 138–144.

421 Valdés, A., Vidal, L., Beltran, A., Canals, A., & Garrigós, M. C. (2015). Microwave-
422 assisted extraction of phenolic compounds from almond skin byproducts (*prunus*
423 *amygdalus*): A multivariate analysis approach. *Journal of Agricultural and Food*
424 *Chemistry*, *63*, 5395–5402.

425 Varela, P., Chen, J., Fiszman, S., & Povey, M. J. W. (2006). Crispness assessment of
426 roasted almonds by an integrated approach to texture description: texture,
427 acoustics, sensory and structure. *Journal of Chemometrics*, *20(6-7)*, 311–320.

428 Vázquez-Araújo, L., Enguix, L., Verdú, A., García-García, E., & Carbonell-Barrachina,
429 A. A. (2008). Investigation of aromatic compounds in toasted almonds used for
430 the manufacture of turrón. *European Food Research and Technology*, *227(1)*,
431 243–254.

- 432 Vlachos, N., Skopelitis, Y., Psaroudaki, M., Konstantinidou, V., Chatzilazarou, A., &
433 Tegou, E. (2006). Applications of Fourier transform-infrared spectroscopy to
434 edible oils. *Analytica Chimica Acta*, 573, 459–465.
- 435 Wold, S., Sjöström, M., & Eriksson, L. (2011). PLS-regression: a basic tool of
436 chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58, 109–30.
- 437 Workman Jr, J., & Weyer, L. (2008). Practical guide to interpretive near-infrared
438 spectroscopy. United States: CRC press, Taylor & Francis.
- 439 Yada, S., Lapsley, K., & Huang, G. (2011). A review of composition studies of
440 cultivated almonds: Macronutrients and micronutrients. *Journal of Food*
441 *Composition and Analysis*, 24(4-5), 469–480.
- 442 Zamany, A. J., Samadi, G. R., Kim, D. H., Keum, Y. S., & Saini, R. K. (2017).
443 Comparative study of tocopherol contents and fatty acids composition in twenty
444 almond cultivars of Afghanistan. *Journal of the American Oil Chemists' Society*,
445 94(6), 805–817.
- 446 Zhang, L. F., Guo, B. H., & Zhang, Z. M. (2002). Synthesis of multifunctional
447 polypropylene via solid phase cografting and its grafting mechanism. *Journal of*
448 *Applied Polymer Science*, 84, 929–935.

449 **Figure captions**

450

451 **Fig. 1.** Representative raw and pre-treated (a) NIR and (b) ATR-FTIR spectra of intact
452 almonds.

453

454 **Fig. 2.** PCA score and X-loading plots of the two first PCs using (a,c) NIR and (b,d)
455 ATR-FTIR pre-treated spectral data, respectively.

456

457 **Fig. 3.** Predicted values for the test set almonds of the PLS-DA models constructed with
458 (a) NIR and (b) ATR-FTIR data.

459

460 **Fig. 4.** QDA plots constructed with (a) NIR and (b) ATR FT-IR data for the
461 discrimination of the test set almonds according to their variety.

Highlights

- Varietal classification of intact Spanish almonds using NIR and ATR-FTIR.
- QDA and PLS-DA were applied to both NIR and ATR-FTIR pre-treated spectral data.
- A performance of 94.45% was obtained with both PLS-DA of ATR-FTIR and QDA of NIR.

Figure 1

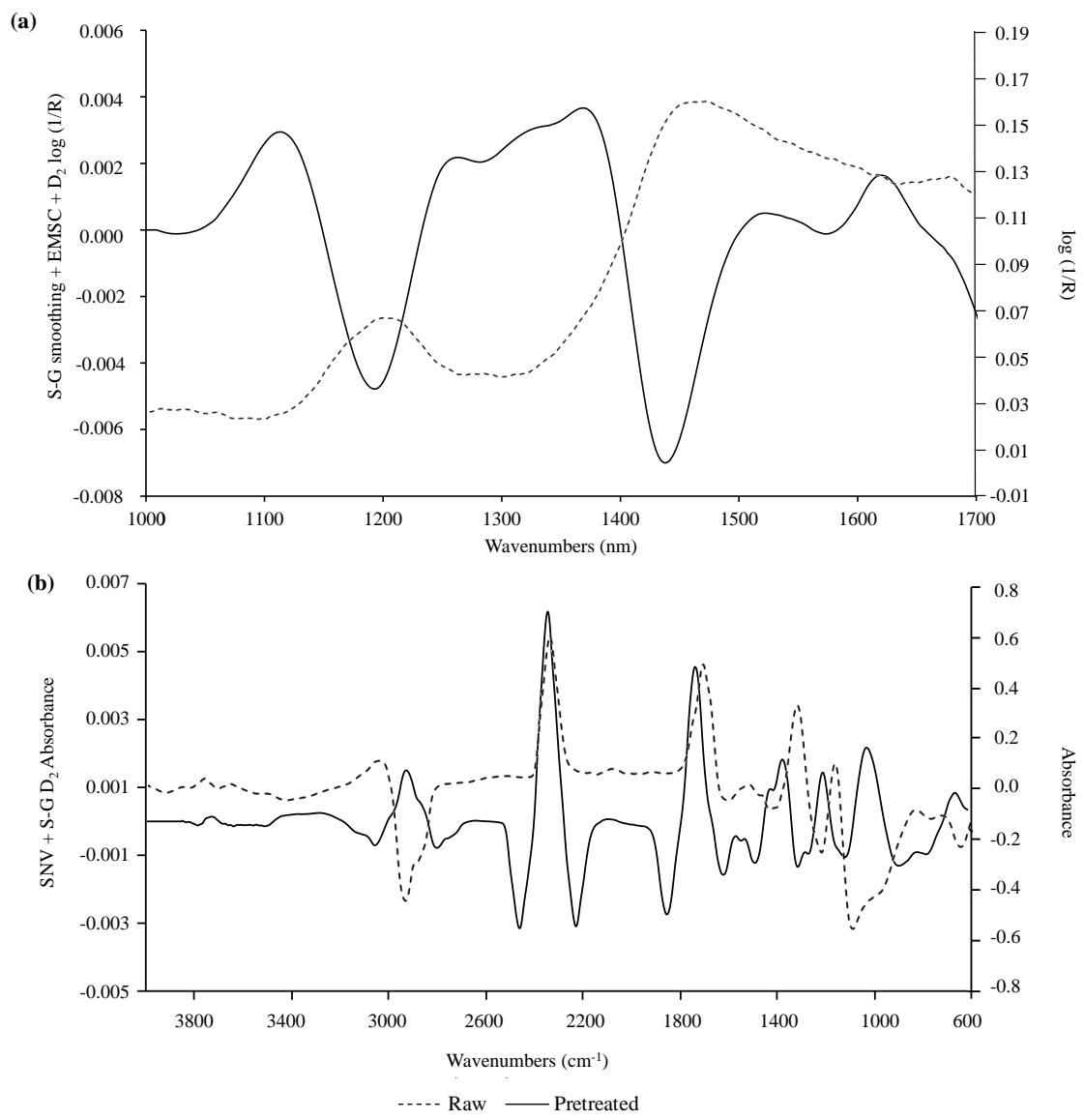


Figure 1. V. Cortés et al.

Figure 2

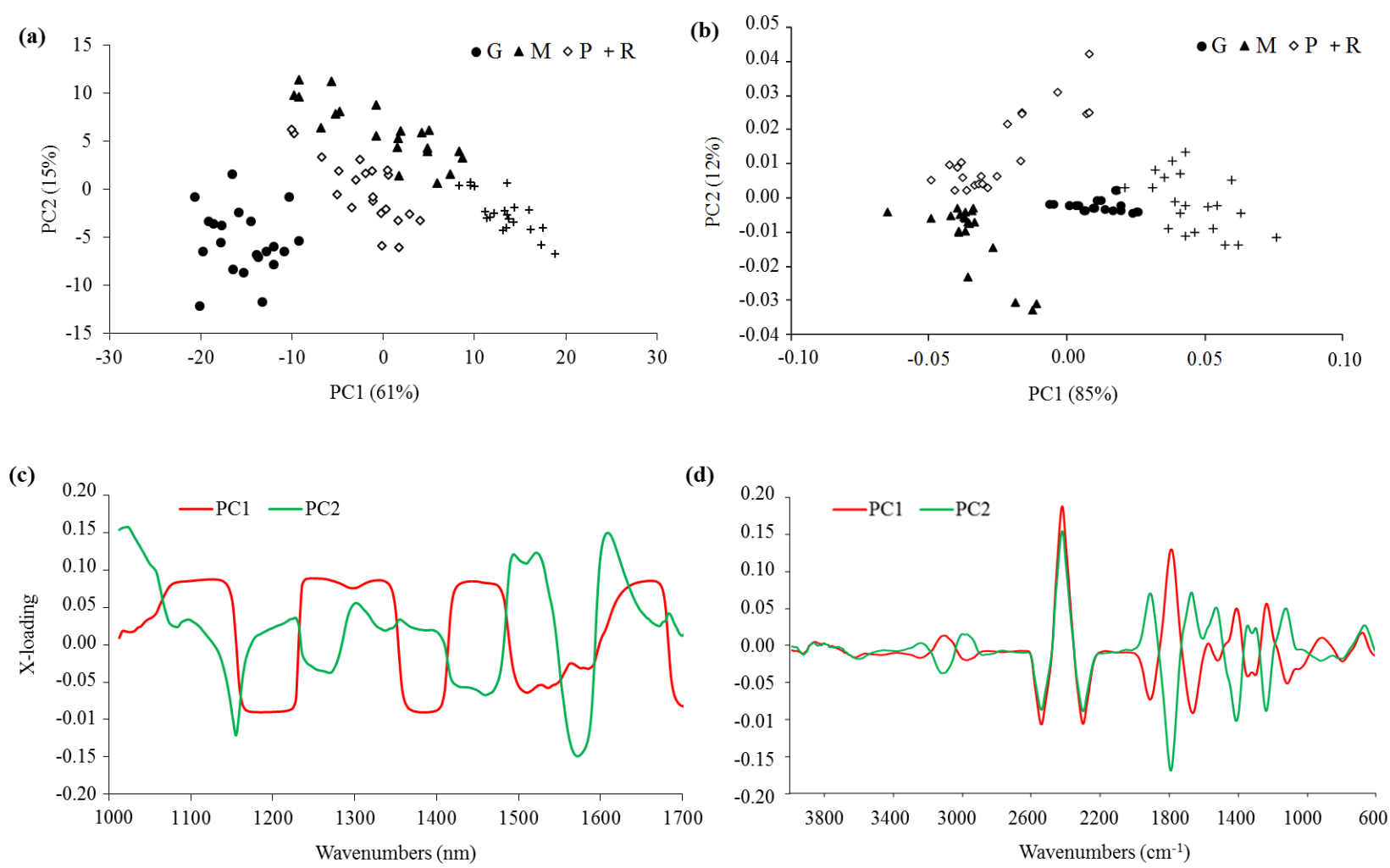


Figure 2. V. Cortés et al.

Figure 3

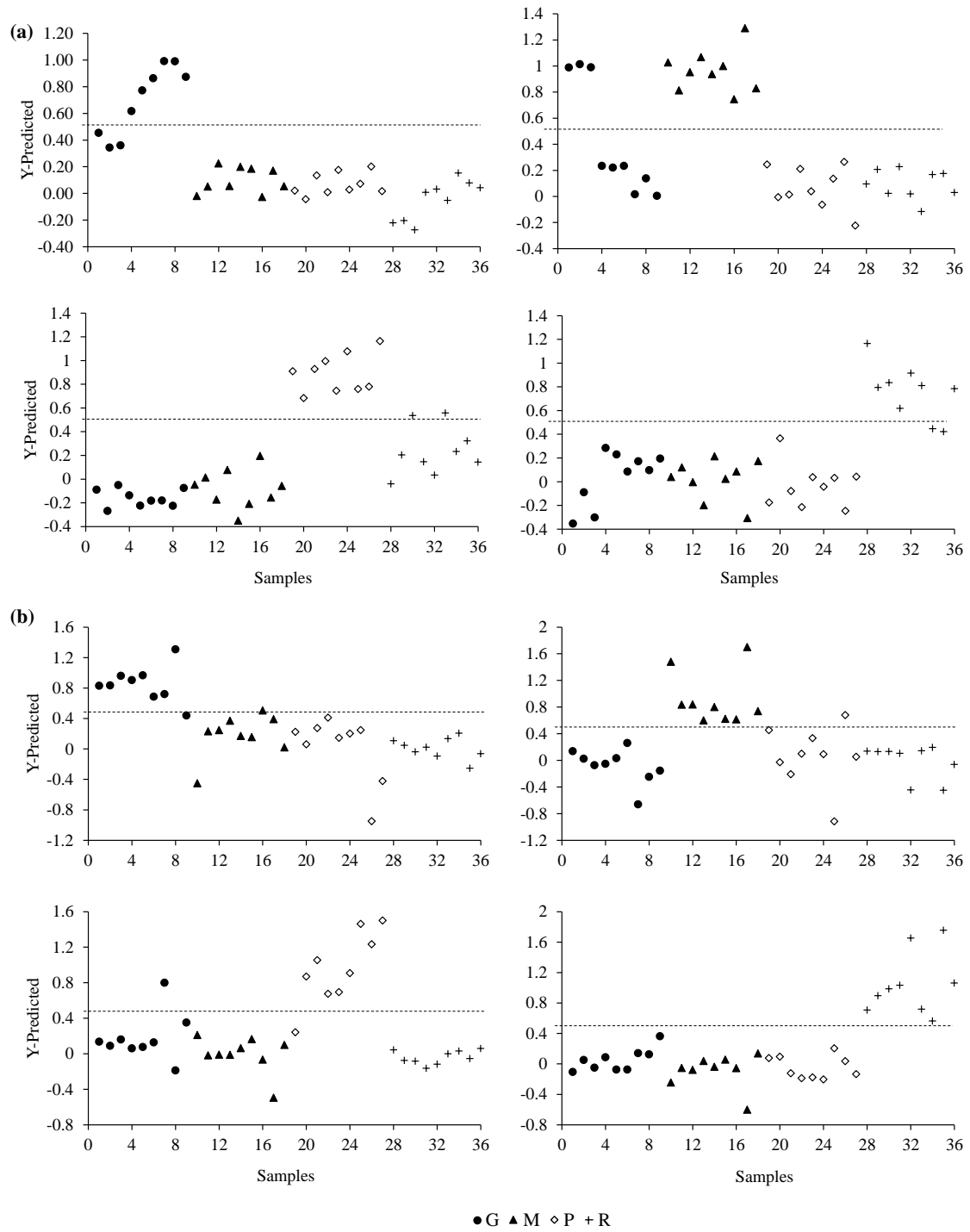


Figure 3. V. Cortés et al.

Figure 4

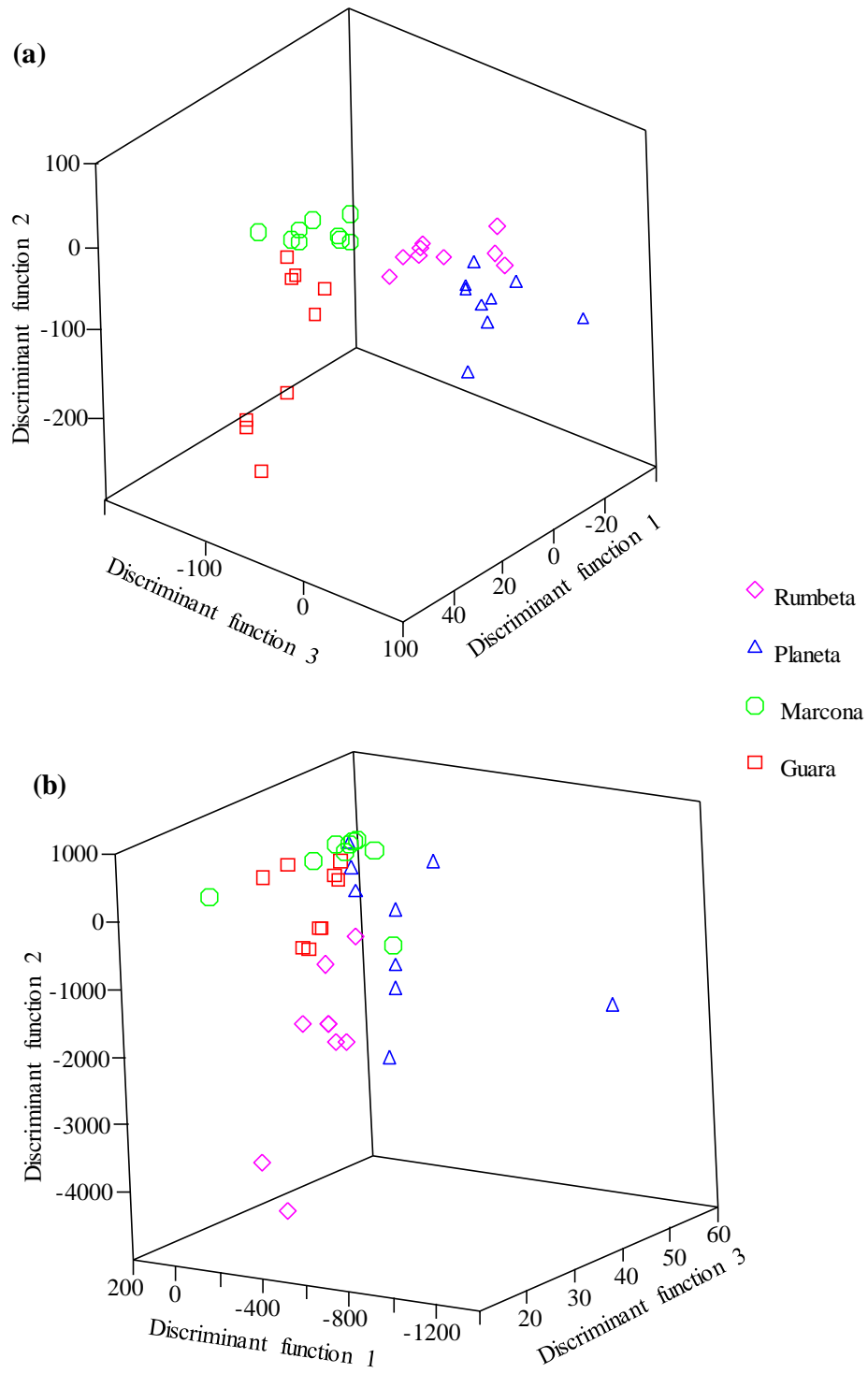


Figure 4. V. Cortés et al.

Table 1

Results of the accuracy of the PLS-DA models constructed to classify almonds according to their variety using training set samples.

	Categories	Calibration		Cross-validation	
		R^2_C	RMSEC	R^2_{CV}	RMSECV
NIR	<i>Guara</i>	0.93	0.10	0.91	0.12
	<i>Marcona</i>	0.94	0.11	0.91	0.13
	<i>Planeta</i>	0.93	0.11	0.90	0.14
	<i>Rumbeta</i>	0.91	0.14	0.85	0.18
ATR-FTIR	<i>Guara</i>	0.94	0.10	0.87	0.15
	<i>Marcona</i>	0.93	0.11	0.86	0.16
	<i>Planeta</i>	0.97	0.08	0.92	0.13
	<i>Rumbeta</i>	0.96	0.09	0.91	0.13

R^2_C = coefficient of determination for calibration; RMSEC = root mean square error of calibration; R^2_{CV} = coefficient of determination for cross-validation; RMSECV = root mean square error of cross-validation.

Table 2

PLS-DA and QDA classification results of test set almond samples using NIR and ATR-FTIR data.

		Correct classification					
		Guara	Marcona	Planeta	Rumbeta	Total (%)	
PLS-DA	NIR	Categories					
		Guara	6/9 (66.7%)	3	0	0	86.13
		Marcona	0	9/9 (100%)	0	0	
		Planeta	0	0	9/9 (100%)	0	
	Rumbeta	0	0	2	7/9 (77.8%)		
	ATR-FTIR	Categories					
		Guara	8/9 (88.9%)	0	1	0	94.45
		Marcona	0	9/9 (100%)	0	0	
Planeta		0	1	8/9 (88.9%)	0		
Rumbeta	0	0	0	9/9 (100%)			
QDA	NIR	Categories					
		Guara	9/9 (100%)	0	0	0	94.45
		Marcona	1	8/9 (88.9%)	0	0	
		Planeta	0	0	8/9 (88.9%)	1	
	Rumbeta	0	0	0	9/9 (100%)		
	ATR-FTIR	Categories					
		Guara	5/9 (55.6%)	2	2	0	77.80
		Marcona	1	7/9 (77.8%)	1	0	
Planeta		1	0	8/9 (88.9%)	0		
Rumbeta	0	0	1	8/9 (88.9%)			